International Journal of Approximate Reasoning 50 (2009) 1115–1128

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Finite approximations of data-based decision problems under imprecise probabilities

Robert Hable *

University of Bayreuth, Department of Mathematics, 95440 Bayreuth, Germany

ARTICLE INFO

Article history:
Received 26 January 2009
Received in revised form 4 May 2009
Accepted 6 May 2009
Available online 20 May 2009

Keywords:
Imprecise probability
Coherent lower prevision
Data-based decision theory
Discretization

ABSTRACT

In decision theory under imprecise probabilities, discretizations are a crucial topic because many applications involve infinite sets whereas most procedures in the theory of imprecise probabilities can only be calculated for finite sets so far. The present paper develops a method for discretizing sample spaces in data-based decision theory under imprecise probabilities. The proposed method turns an original decision problem into a discretized decision problem. It is shown that any solution of the discretized decision problem approximately solves the original problem.

In doing so, it is pointed out that the commonly used method of natural extension can be most instable. A way to avoid this instability is presented which is sufficient for the purpose of the paper.

1. Introduction

1.1. Motivation

Decision theory provides a formal framework which covers many potential areas of application including e.g. economics or statistics. As a matter of fact, most statistical problems can be formulated as decision problems. This enables a more unified treatment of statistical problems and has become an important area in statistics called statistical decision theory; cf. e.g. [1–3].

If we are faced with a decision problem, we have to choose a decision and every decision t leads to a certain loss $W_\theta(t)$ which depends on an unknown state of nature $\theta \in \Theta$. Most often, the decision can be made on base of an observation $x \in \mathscr{X}$. This observation x is distributed according to some probability measure P_θ which is known except for the unknown θ. In a statistical problem for example, $(P_\theta)_{\theta\in\Theta}$ is a parametric model, x represents the data and a decision t might be an estimation of the true θ or the decision whether a hypothesis should be rejected or not. However, assuming a precisely known model $(P_\theta)_{\theta\in\Theta}$ which consists of probability measures is often unrealistic in real applications since the model $(P_\theta)_{\theta\in\Theta}$ usually will only imprecisely be known. This can be formulated by replacing $(P_\theta)_{\theta\in\Theta}$ by an "imprecise model" $(\overline{P}_\theta)_{\theta\in\Theta}$ where $\overline{P}_\theta$ is a suitable generalization of a (precise) probability measure. Such generalizations have recently been developed among others, by Walley [4] and Weichselberger [5]. Here, the probability of an event is no longer a precise number $p \in [0,1]$ but an interval $[\underline{p}, \overline{p}] \subset [0,1]$. Concepts of imprecise probabilities have already been used in many decision theoretic investigations. For example, a classical text in mathematical economics is [6]. General articles about decision making under imprecise probabilities are [7,8].

* Tel.: +49 0 921/55 3274; fax: +49 0 921/55 2152.
E-mail address: Robert.Hable@uni-bayreuth.de

0888-613X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2009.05.003

There are a number of algorithms available now which make applications possible; cf. e.g. [9,10] and, in particular for decision theoretic purposes, [11,12]. However, these algorithms are based on linear programming and, therefore, can only deal with finite sets. Since this is a severe restriction for applications, discretization is a crucial topic in the theory of imprecise probabilities which has recently been considered in Troffaes [13] and Obermeier and Augustin [14]. These two papers consider decision theory which is not explicitly data-based and, accordingly, are concerned with discretizations of the set Θ of all states of nature. For precise models and precise prior distributions, the often so-called 'main theorem of Bayesian decision theory' [2, p. 159] implies that a data-based decision problem may be solved by considering a corresponding data-free decision problem which is generated by updating the precise prior distribution. However, according to [15], updating imprecise prior distributions, in general, leads to suboptimal results so that data-based decision theory can be seen as a matter of its own in case of imprecise probabilities.

In data-based decision theory under imprecise probabilities, not only discretizing Θ but also discretizing the sample space $(\mathscr{X},\mathscr{A})$ is an important issue – the more so as, e.g. in statistics, infinite sample spaces are at least as common as finite sample spaces. Therefore, it is the purpose of the present paper to develop a method for discretizing the sample space $(\mathscr{X},\mathscr{A})$ which is justified by theoretical results and appropriate to applications.

The setup concerning data-based decision theory under imprecise probabilities is presented in the following subsection. Here, imprecise probabilities are modeled via coherent upper previsions according to [4]. A recent survey of the theory of coherent upper (or lower) previsions is [16]. In addition, it is assumed that a practitioner is only able to specify a finite number of upper previsions $\overline{P}_\theta[f_1], \overline{P}_\theta[f_2], \ldots, \overline{P}_\theta[f_n]$ for each $\theta \in \Theta$. Though this assumption is restrictive, it is very often fulfilled – especially in real applications. In particular, this is true for expert systems since, there, it is a natural approach to ask some experts about their prevision (or expectation) on some specific events, experiments, gambles, assets etc. and this can only be done for a finite number of such objects. In doing so, the upper prevision has to be extended from a finite set of functions to the whole sample space by the method of natural extension, which is quite common in the theory of imprecise probability according to [4].

However, such an approach needs some care and a thoughtless application may lead to arbitrary results because natural extensions are potentially most instable as shown in Section 2. Theorem 2.2 provides a guideline how to avoid such instabilities. Though this result is certainly not fully satisfactory for all applications, it is appropriate at least for the applications which the present paper focuses on. Section 3 is concerned with a special method for discretizing the sample space. After specifying the assumptions and the model in Section 3.1, the proposed method for discretizing is presented in Section 3.2. In this way, the original decision problem is turned into a discretized decision problem. Next, it is proven in Section 3.3 that the original decision problem can approximately be solved by solving the discretized decision problem. Finally, Section 4 discusses the applicability of the proposed method by means of the size of the discretized decision problem. Section 5 contains some concluding remarks.

1.2. Setup

Let $(\mathscr{X},\mathscr{A})$ be a measurable space, i.e. $\mathscr{X}$ is a set and $\mathscr{A}$ is a σ-algebra on $\mathscr{X}$. Let $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ be the set of all bounded, $\mathscr{A}$-measurable functions $f:\mathscr{X}\to\mathbb{R}$ and let $\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})$ denote the set of all finitely additive probability measures on $(\mathscr{X},\mathscr{A})$. That is, $\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})$ is the set of all set functions $P:\mathscr{A}\to[0,1]$ such that $P(\emptyset)=0$, $P(\mathscr{X})=1$ and

$$A_1,\ldots,A_n\in\mathscr{A},\ \ A_i\cap A_j=\emptyset\ \ \forall i\neq j\ \ \Rightarrow\ \ P\left(\bigcup_{i=1}^n A_i\right)=\sum_{i=1}^n P(A_i)$$

Finitely additive probability measures are also called probability charges in Bhaskara Rao and Bhaskara Rao [17]. For every $f\in\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$, integrals are defined with respect to any $P\in\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})$ – see e.g. [18, Section III] or [17, Section 4]. This integral is denoted by

$$P[f]:=\int f dP$$

In this way, P represents a bounded linear operator on $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$, which is called linear prevision in Walley [4].

Let $(\overline{P}_\theta)_{\theta\in\Theta}$ be an imprecise model on $(\mathscr{X},\mathscr{A})$. That is, Θ is an index set and, for every $\theta\in\Theta$,

$$\overline{P}_\theta:\mathscr{L}_\infty(\mathscr{X},\mathscr{A})\mapsto\mathbb{R},\quad f\mapsto\overline{P}_\theta[f]$$

is a coherent upper prevision according to [4, Section 2.5.1]. Based on behavioral interpretations, coherent upper previsions have originally been defined as infimum selling prices in Walley [4]. Equivalently, coherent upper previsions can also be defined in the following way which corresponds to a sensitivity analysis interpretation (see [4, p. 53 and Section 3.3.3] and [19, Section 2.3]): $\overline{P}_\theta$ is a coherent upper prevision on $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ if and only if there is a set $\mathscr{V}_\theta\subset\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})$ such that $\sup_{P_\theta\in\mathscr{V}}P_\theta[f]=\overline{P}_\theta[f]$ for every $f\in\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$. For every $\theta\in\Theta$, the non-empty set

$$\mathscr{M}_\theta=\left\{P_\theta\in\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})\quad\middle|P_\theta[f]\leqslant\overline{P}_\theta[f]\quad\forall f\in\mathscr{L}_\infty(\mathscr{X},\mathscr{A})\right\}$$

is called *credal set of* $\overline{P}_\theta$. Of course, $\sup_{P_\theta\in\mathscr{M}_\theta}P_\theta[f]=\overline{P}_\theta[f]$ then. In other words, $\overline{P}_\theta$ is the upper envelope of a class of linear previsions.

A decision space is a measurable space $(\mathbb{D},\mathscr{D})$ where every element $t \in \mathbb{D}$ represents a specific decision in the decision problem. Accordingly, $\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$ denotes the set of all bounded, $\mathscr{D}$-measurable functions $h:\mathbb{D}\to\mathbb{R}$. An ($\mathscr{A}$-measurable) decision function is a map $\delta:\mathscr{X}\to\mathbb{D}$ which is measurable with respect to $\mathscr{A}$ and $\mathscr{D}$. The decision theoretic interpretation of such a map is: in case of observing $x\in\mathscr{X}$, choose decision $\delta(x)=t\in\mathbb{D}$. An ($\mathscr{A}$-measurable) *randomized* decision function is an $\mathscr{A}$-measurable Markov kernel

$$\tau:\mathscr{X}\times\mathscr{D}\to[0,1],\quad (x,D)\mapsto\tau_x(D)$$

That is, τ_x is a probability measure on $(\mathbb{D},\mathscr{D})$ for every $x\in\mathscr{X}$ and the map $x\mapsto\tau_x(D)$ is an element of $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ for every $D\in\mathscr{D}$. The decision theoretic interpretation is as follows: in case of observing $x\in\mathscr{X}$, start an auxiliary random experiment according to τ_x and choose that decision $t\in\mathbb{D}$ which is the outcome of the auxiliary random experiment. In this way, a decision is made "by coin tossing" where τ_x gives the probabilities with which decisions are selected. Note that these probabilities are precise; this is justified because such auxiliary random experiments correspond to ideal game situations.

The set of all $\mathscr{A}$-measurable decision functions is denoted by $\mathscr{T}_d(\mathscr{A})$ and the set of all $\mathscr{A}$-measurable randomized decision functions is denoted by $\mathscr{T}_0(\mathscr{A})$. Since every decision function δ defines a randomized decision function

$$\tau:(x,D)\mapsto\tau_x(D)=\begin{cases}1 & \text{if } \delta(x)\in D\\ 0 & \text{if } \delta(x)\notin D\end{cases},$$

$\mathscr{T}_d(\mathscr{A})$ may be considered as a subset of $\mathscr{T}_0(\mathscr{A})$.

The elements θ of the index set Θ represent the possible states of nature. It is assumed that there is a true but unknown state of nature $\theta_0\in\Theta$. A loss function is a map $W:\Theta\times\mathbb{D}\to\mathbb{R}$, $(\theta,t)\mapsto W_\theta(t)$ such that $(W_\theta)_{\theta\in\Theta}\subset\mathscr{L}_\infty(\mathbb{D},\mathscr{D})$. The real number $W_\theta(t)$ represents the loss which occurs if θ is the true state of nature and t is our decision.

With respect to such a loss function and the imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$, the (upper) risk function of a randomized decision function $\tau\in\mathscr{T}_0(\mathscr{A})$ is denoted by

$$R\big((\overline{P}_\theta)_{\theta\in\Theta},\tau,W\big):\Theta\to\mathbb{R},\quad \theta\mapsto R\big(\overline{P}_\theta,\tau,W_\theta\big)$$

where, for every $\theta\in\Theta$, $R\big(\overline{P}_\theta,\tau,W_\theta\big)$ is defined to be

$$R\big(\overline{P}_\theta,\tau,W_\theta\big)=\overline{P}_\theta[\tau_\bullet[W_\theta]]=\sup_{P_\theta\in\mathscr{M}_\theta}\int_{\mathscr{X}}\int_{\mathbb{D}}W_\theta(t)\tau_x(dt)P_\theta(dx)$$

That is, the value of the risk function is the supremal expected losses of a (randomized) decision function depending on the state of nature.

For an introduction to such decision theoretic concepts with precise probabilities, see [2]; For a more detailed description of (data-based) decision theory under imprecise probabilities, see e.g. [20].

As already mentioned in Section 1.1, we restrict our considerations to coherent upper previsions which are natural extensions of coherent upper previsions on finite sets.

That is, we assume that, for every $\theta\in\Theta$, there is a *finite* subset $\mathscr{K}_\theta\subset\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ such that the credal set of $\overline{P}_\theta$ is given by

$$\mathscr{M}_\theta=\big\{P_\theta\in\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})\big|P_\theta[f]\leqslant\overline{P}_\theta[f]\forall f\in\mathscr{K}_\theta\big\}\tag{1}$$

On the one hand, these coherent upper previsions have particular importance in applications since often only a finite number of upper previsions $\overline{P}_\theta[f]$ can be explicitly specified by practitioners. On the other hand, these models needs some care since the method of natural extension can be most instable. Therefore, the following section takes a closer look on natural extensions in this model.

2. Instability of natural extensions

According to Section 1.2, we are concerned with the following way of modeling: it is assumed that a practitioner is only able to explicitly specify the coherent upper prevision on a finite subset $\mathscr{K}\subset\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ so that we are faced with a coherent upper prevision

$$\overline{P}:\mathscr{K}\to\mathbb{R},\quad f\mapsto\overline{P}[f]$$

It is one of the prime characteristics of the theory of imprecise probabilities that this – at least in theory – does not provide any problem because $\overline{P}$ can always coherently be extended to a coherent upper prevision

$$\overline{P}:\mathscr{L}_\infty(\mathscr{X},\mathscr{A})\to\mathbb{R},\quad f\mapsto\overline{P}[f]$$

The credal set of the extended coherent upper prevision is given by

$$\mathscr{M}=\big\{P\in\mathrm{ba}_1^+(\mathscr{X},\mathscr{A})\big|P[f]\leqslant\overline{P}[f]\forall f\in\mathscr{K}\big\}$$

However, severe problems may arise in practical applications:

The main justification of developing a theory of imprecise probability is the fact that it is usually not possible to specify correct precise probabilities. Of course, it is far more easy and realistic to determine upper and lower bounds for the probabilities than to determine precise probabilities. However, it is also unrealistic to assume that practitioners can precisely specify "correct" upper and lower bounds for the probabilities and, therefore, small changes in the upper and lower bounds should only have small effects in the evaluation. This will usually be true but, unfortunately, this is not always true. Arbitrarily small changes in the upper and lower bounds can have arbitrarily large effects in the above model.

To demonstrate this, assume that there is another coherent upper prevision

$$\overline{P}' : \mathscr{K} \to \mathbb{R}, \quad f \mapsto \overline{P}'[f]$$

and that there is some $f_0 \in \mathscr{K}$ such that

$$\overline{P}[f] = \overline{P}'[f] \quad \forall f \in \mathscr{K} \setminus \{f_0\} \quad \text{and} \quad \left|\overline{P}[f] - \overline{P}'[f]\right| \leqslant \varepsilon \in (0,1)$$

Then, the worst case would be if the following was possible for the natural extensions – regardless how small $\varepsilon > 0$ is:

$$\overline{P}[f] = \inf f \quad \text{but} \quad \overline{P}'[f] = \sup f$$

for some (non-constant) $f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$. The following simple example shows that this worst case, in fact, may happen:

Example 2.1. Take $\mathscr{X} = [0,1]$ and let $\mathscr{A}$ be the Borel-σ-algebra of [0,1]. Define $f_0 : [0,1] \to \mathbb{R}$, $f_0(x) = x \ \forall x \in [0,1]$ and $\mathscr{K} = \{f_0\}$. Furthermore,

$$\overline{P}[f_0] = 0 \quad \text{and} \quad \overline{P}'[f_0] = \varepsilon$$

where $0 < \varepsilon < 1$. Then, the credal sets of the natural extensions are given by

$$\mathscr{M} = \left\{P \in \mathrm{ba}_1^+(\mathscr{X}, \mathscr{A}) | P[f_0] = 0\right\}$$

and

$$\mathscr{M}' = \left\{P' \in \mathrm{ba}_1^+(\mathscr{X}, \mathscr{A}) \middle| P'[f_0] \in [0, \varepsilon]\right\}$$

It follows from $0 \leqslant \varepsilon I_{[\varepsilon,1]} \leqslant f_0$ that $\overline{P}\left[\varepsilon I_{[\varepsilon,1]}\right] = 0$. Hence, $\overline{P}\left[I_{[\varepsilon,1]}\right] = 0$.

Let δ_ε be the Dirac measure in ε. Then, $\delta_\varepsilon[f_0] = \varepsilon$ implies $\delta_\varepsilon \in \mathscr{M}'$ and, together with $I_{[\varepsilon,1]} \leqslant 1$, this implies $\overline{P}'\left[I_{[\varepsilon,1]}\right] = 1$. Summing up, we have

$$\overline{P}\left[I_{[\varepsilon,1]}\right] = \inf_{x \in [0,1]} I_{[\varepsilon,1]}(x) \quad \text{but} \quad \overline{P}'\left[I_{[\varepsilon,1]}\right] = \sup_{x \in [0,1]} I_{[\varepsilon,1]}(x)$$

Note that the above example is not a pathological one: the sample space is a compact interval in $\mathbb{R}$, the σ-algebra $\mathscr{A}$ is the Borel-σ-algebra and the coherent upper prevision on $\mathscr{K}$ is a very easy one because $\mathscr{K}$ only consists of one element f_0 and this f_0 is a linear function. However, the above example, indeed, is somehow special because we have $\underline{P}[f_0] = \overline{P}[f_0]$ and this is a precise prevision which is not really what we want in imprecise probabilities. Nevertheless, the use of such imprecise probabilities where

$$\underline{P}[f] = \overline{P}[f] \tag{2}$$

at least for some (non-constant) functions $f \in \mathscr{K}$ is not unusual in applications of imprecise probabilities.

For applications, it would be desirable to have some guidelines which prevent practitioners from arbitrary results because of an instable natural extension. The following theorem makes a first attempt in this direction. Though this is sufficient within the scope of the present article concerning discretizing, the theorem certainly does not succeed in giving a final, satisfactory answer. Hopefully, future research will provide some more insight into this important topic.

Theorem 2.2. *Let $\mathscr{K}$ be any subset of $\mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$ and let*

$$\mathscr{F} := \{f_1, \ldots, f_n\} \subset \mathscr{K} \subset \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$$

be a finite subset of $\mathscr{K}$.

Let $\overline{P}$ and $\overline{P}'$ be coherent upper previsions on $\mathscr{K}$ such that

$$\overline{P}[f] = \overline{P}'[f] \quad \forall f \in \mathscr{K} \setminus \mathscr{F}$$

and, for some real numbers $0 < \varepsilon_i < 1$, $i \in \{1, \ldots, n\}$,

$$\overline{P}[f_i] \leqslant \overline{P}'[f_i] \leqslant \overline{P}[f_i] + \varepsilon_i\left(\overline{P}[f_i] - \underline{P}[f_i]\right) \quad \forall i \in \{1, \ldots, n\} \tag{3}$$

where $\underline{P}$ is the coherent lower prevision on $\mathscr{K}$ which corresponds to $\overline{P}$. [1]

Let $\underline{P}$, $\overline{P}$ and $\overline{P}'$ also denote the respective natural extensions of $\underline{P}$, $\overline{P}$ and $\overline{P}'$ on $\mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$. Then, for $\epsilon := \varepsilon_1 + \ldots + \varepsilon_n$:

[1] $\underline{P}[f] = \inf_{P \in \mathscr{M}} P[f] \ \forall f \in \mathscr{K}$ *where $\mathscr{M}$ is the credal set of $\overline{P}$.*

$$\overline{P}[f] \leqslant \overline{P}'[f] \leqslant \overline{P}[f] + \varepsilon(\sup f - \underline{P}[f]) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$$

That is, Theorem 2.2 investigates what may happen (or rather what cannot happen) if some values of a coherent upper prevision $\overline{P}$ are made slightly larger. "Slightly" means: a small percentage of $\overline{P}[f] - \underline{P}[f]$. Therefore, $\overline{P}[f]$ may not be changed if $\overline{P}[f] = \underline{P}[f]$ so that the above example is excluded. So, Theorem 2.2 explains how to avoid instability of the natural extension by avoiding such bottlenecks (2). Theorem 2.2 is a consequence of the following lemma, which is interesting on its own.

Lemma 2.3. *Under the assumptions of Theorem 2.2, let $\mathscr{M}$ be the credal set of $\overline{P}$ and $\mathscr{M}'$ the credal set of $\overline{P}'$. Then, $\mathscr{M} \subset \mathscr{M}'$ and, for every $P' \in \mathscr{M}'$, there is a $P \in \mathscr{M}$ such that*

$$P'[f] \leqslant P[f] + \varepsilon(\sup f - \underline{P}[f]) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A}) \tag{4}$$

and

$$\left|P[f] - P'[f]\right| \leqslant \varepsilon(\sup f - \inf f) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A}) \tag{5}$$

Proof of Lemma 2.3: The assumptions immediately imply

$$\mathscr{M} \subset \mathscr{M}' \quad \text{hence} \quad \overline{P}'[f] \geqslant \underline{P}[f] \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A}) \tag{6}$$

Fix any $P' \in \mathscr{M}'$. For every $i \in \{1, \ldots, n\}$, there is a $P_i \in \mathscr{M}$ such that $\underline{P}[f_i] = P_i[f_i]$; cf. [4, Section 3.6.2]. Define $P'_0 := P'$ and consider the following inductive definitions for $i \in \{1, \ldots, n\}$:

- In case of $P'_{i-1}[f_i] \leqslant \overline{P}[f_i]$ (CASE 1), define $\alpha_i = 0$.
- In case of $P'_{i-1}[f_i] > \overline{P}[f_i]$ (CASE 2), define.

$$\alpha_i := \frac{P'_{i-1}[f_i] - \overline{P}[f_i]}{P'_{i-1}[f_i] - P_i[f_i]}$$

Then, define $P'_i := (1 - \alpha_i)P'_{i-1} + \alpha_i P_i$.

Now, we want to show that (4) is fulfilled for $P = P'_n$. To this end, fix any $f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$. By induction, we prove in the following that, for every $i \in \{0, \ldots, n\}$,

$$P'_i \in \mathscr{M}' \tag{7}$$

$$P'_i[f_j] \leqslant \overline{P}[f_j] \quad \forall j \in \{1, \ldots, i\} \tag{8}$$

and

$$P'[f] \leqslant P'_i[f] + \sum_{j=1}^{i} \varepsilon_j \cdot \left(\overline{P}'[f] - \underline{P}[f]\right) \tag{9}$$

Obviously, (7)–(9) are fulfilled for $i = 0$. Next, let (7)–(9) be fulfilled for $i - 1$.

– *CASE 1*: In case of $P'_{i-1}[f_i] \leqslant \overline{P}[f_i]$, we have $P_i = P_{i-1}$ and, therefore, it is easy to see that (7)–(9) are fulfilled.
– *CASE 2*: In case of $P'_{i-1}[f_i] > \overline{P}[f_i]$, it follows from

$$\underline{P}[f_i] = P_i[f_i] \leqslant \overline{P}[f_i] < P'_{i-1}[f_i] \overset{(7)}{\leqslant} \overline{P}[f_i] + \epsilon_i\left(\overline{P}[f_i] - \underline{P}[f_i]\right)$$

that

$$0 \leqslant \alpha_i = \frac{P'_{i-1}[f_i] - \overline{P}[f_i]}{P'_{i-1}[f_i] - P_i[f_i]} \leqslant \frac{\varepsilon_i(\overline{P}[f_i] - \underline{P}[f_i])}{\overline{P}[f_i] - \underline{P}[f_i]} = \varepsilon_i \tag{10}$$

then. In particular, $\alpha_i \in [0, 1]$. Next, the definition of P'_i, (6) and the induction hypothesis immediately imply the validity of (7) for i and

$$P'_i[f_j] \leqslant \overline{P}[f_j] \quad \forall j \in \{1, \ldots, i - 1\}$$

Furthermore,

$$P'_i[f_i] = (1 - \alpha_i)P'_{i-1}[f_i] + \alpha_i P_i[f_i] = \frac{\overline{P}[f_i] - P_i[f_i]}{P'_{i-1}[f_i] - P_i[f_i]} \cdot P'_{i-1}[f_i] + \frac{P'_{i-1}[f_i] - \overline{P}[f_i]}{P'_{i-1}[f_i] - P_i[f_i]} \cdot P_i[f_i] = \overline{P}[f_i]$$

That is, we have proven the validity of (7) and (8) for i so far. In order to prove (9), note that $(1 - \alpha_i)P'_{i-1} + \alpha_i P_i = P'_i$ implies

$$P'_{i-1}[f] = \alpha_i P'_{i-1}[f] + P'_i[f] - \alpha_i P_i[f] \overset{(7)}{\leqslant} P'_i[f] + \alpha_i\left(\overline{P}'[f] - \underline{P}[f]\right) \leqslant P'_i[f] + \varepsilon_i\left(\overline{P}'[f] - \underline{P}[f]\right)$$

where the last inequality follows from (6) and (10). Together with the induction hypothesis, this implies

$$P'[f] \leqslant P'_{i-1}[f] + \sum_{j=1}^{i-1} \varepsilon_j \cdot \left(\overline{P}'[f] - \underline{P}[f]\right) \leqslant P'_i[f] + \sum_{j=1}^{i} \varepsilon_j \cdot \left(\overline{P}'[f] - \underline{P}[f]\right)$$

Hence, the inductive step is also proven for CASE 2 now. Summing up, we have proven by induction the validity of (7)–(9) for every $i \in \{1, \ldots, n\}$ so far.

For $i = n$, (7) and (8) imply $P = P'_n \in \mathscr{M}$ and (4) follows from (9). Finally, (4) implies $P'[f] \leqslant P[f] + \varepsilon(\sup f - \inf f)$ and

$$P'[f] = -P'[-f] \geqslant -P[-f] - \varepsilon(\sup(-f) - \inf(-f)) = P[f] - \varepsilon(\sup f - \inf f)$$

for every $f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$. Hence, (5) follows. □

Proof of Theorem 2.2. This is an immediate consequence of assertion (4) of Lemma 2.3. □

Remark 2.4. Of course, Proposition 2.2 can be simplified to the weaker bound

$$\overline{P}[f] \leqslant \overline{P}'[f] \leqslant \overline{P}[f] + \varepsilon(\sup f - \inf f) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$$

3. A method for discretizing sample spaces

3.1. Assumptions and modeling

According to the setup presented in Section 1.2, we have an imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on $(\mathscr{X}, \mathscr{A})$. The proposed method for discretizing the sample space $(\mathscr{X}, \mathscr{A})$ needs the following three assumptions:

It is assumed that, for every $\theta \in \Theta$, there is a <u>finite</u> subset $\mathscr{K}_\theta \subset \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$ such that the credal set of $\overline{P}_\theta$ is given by

$$\mathscr{M}_\theta = \left\{P_\theta \in \mathrm{ba}_1^+(\mathscr{X}, \mathscr{A}) \middle| P_\theta[f] \leqslant \overline{P}_\theta[f] \ \ \forall f \in \mathscr{K}_\theta\right\} \tag{11}$$

Furthermore, it is assumed that

$$\mathscr{K} := \bigcup_{\theta\in\Theta} \mathscr{K}_\theta \quad \text{is a finite set} \tag{12}$$

Finally, it is assumed, that, for every fixed $f \in \mathscr{K}$, there is a $d_f > 0$ such that

$$\overline{P}_\theta[f] - \underline{P}_\theta[f] \geqslant d_f \quad \text{for every} \quad \theta \in \Theta \text{ where } \mathscr{K}_\theta \ni f \tag{13}$$

Assumption (11) is crucial and rather restrictive – nevertheless, such imprecise models are quite important for practical applications. The index set Θ is not assumed to be finite here. Instead, the considerably weaker Assumption (12) is sufficient. Let Assumption (11) be fulfilled; then (12) is fulfilled if Θ is finite but it is also fulfilled if $\mathscr{K}_\theta$ does not depend on $\theta \in \Theta$. If Θ is finite, then Assumption (13) coincides with the assumption

$$\overline{P}_\theta[f] \neq \underline{P}_\theta[f] \quad \forall f \in \mathscr{K}_\theta \tag{14}$$

Adding Assumption (14) is not restrictive at all since Example 2.1 and Theorem 2.2 tell us: Using models of form (11) which violate (14) is dangerous because these models are potentially most instable. Therefore, those models which violate (14) generally should be avoided anyway.

For practical applications, it is important that the validity of these assumptions can easily be checked: usually, the validity of (11) and (12) directly results from modeling: a practitioner specifies concrete upper previsions for a finite number of functions $f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$ in order to get coherent upper previsions

$$\overline{P}_\theta : \mathscr{K}_\theta \to \mathbb{R}, \quad \theta \in \Theta \tag{15}$$

Next, these coherent upper previsions are extended on $\mathscr{L}_\infty(\mathscr{X}, \mathscr{A})$ by the method of natural extension and this leads to an imprecise model which fulfills (11) and (12). If Θ is infinite, then the infinite number of upper previsions in (15) has been specified by analytical arguments and, therefore, the validity of (13) has to be checked also by analytical arguments. If Θ is finite, then the following procedure can be applied in order to check (13):

For every $f_0 \in \mathscr{K}$, take a partition $\{B_1, \ldots, B_k\}$ of $\mathscr{X}$, fix any $b_j \in B_j \ \forall j \in \{1, \ldots, k\}$ and define

$$\bar{f} := \sum_{j=1}^{k} \sup_{x_j\in B_j} f(x_j) \cdot I_{B_j} \quad \forall f \in \mathscr{K}$$

For every $\theta \in \Theta$ such that $\mathscr{K}_\theta \ni f_0$, solve the following linear program:

$$\left(-\bar{f}_0(b_1), \ldots, -\bar{f}_0(b_k)\right) \cdot p \to \max_p \tag{16}$$

where

$$\left(\bar{f}(b_1),\ldots,\bar{f}(b_k)\right)\cdot p \leqslant \overline{P}_\theta[f] \quad \forall f \in \mathscr{K}_\theta$$

and

$$p \in \mathbb{R}^k, \quad p_j \geqslant 0 \quad \forall j \in \{1,\ldots,k\}, \quad p_1+\ldots+p_k = 1$$

Every feasible solution p of this linear program represents a probability charge on the partition $\{B_1,\ldots,B_k\}$ and the set of all feasible solutions corresponds to the credal set of a coherent upper prevision which approximates $\overline{P}_\theta$.

If the optimal value $l_{f_0,\theta}$ is not larger than $-\overline{P}_\theta[f]$, start again with a finer partition. If the optimal values $l_{f_0,\theta}$ are larger than $-\overline{P}_\theta[f_0]$ for every $\theta \in \Theta$ such that $\mathscr{K}_\theta \ni f$, define

$$d_{f_0} := \min\left\{\overline{P}_\theta[f_0] + l_{f_0,\theta} \middle| \quad \theta \in \Theta : \mathscr{K}_\theta \ni f_0\right\}$$

Assumption (11) is fulfilled, if this procedure ends up with positive numbers $d_f, f \in \mathscr{K}$. This is a consequence of the following proposition which states that $\overline{P}_\theta[f] + l_{f,\theta}$ is a lower bound on $\overline{P}_\theta[f] - \underline{P}_\theta[f]$. Of course, the finer partition $\{B_1,\ldots,B_l\}$ is, the better lower bound $\overline{P}_\theta[f] + l_{f,\theta}$ usually is.

Proposition 3.1. *For a fixed $f_0 \in \mathscr{K}$ and a fixed $\theta \in \Theta$, assume that the linear program (16) has an optimal value $l_{f_0,\theta}$. Then,*

$$\overline{P}_\theta[f_0] + l_{f_0,\theta} \leqslant \overline{P}_\theta[f_0] - \underline{P}_\theta[f_0]$$

Proof. Take

$$\widehat{\mathscr{M}}_\theta := \left\{P_\theta \in \mathrm{ba}_1^+(\mathscr{X},\mathscr{A}) \middle| \quad P_\theta\left[\bar{f}\right] \leqslant \overline{P}_\theta[f] \quad \forall f \in \mathscr{K}_\theta\right\}$$

The construction implies that the optimal value $l_{f_0,\theta}$ in the linear program (16) is equal to

$$l_{f_0,\theta} = \sup_{P_\theta \in \widehat{\mathscr{M}}_\theta} P_\theta\left[-\bar{f}_0\right]$$

Note that $f \leqslant \bar{f}$ for every $f \in \mathscr{K}_\theta$. Hence, $\widehat{\mathscr{M}}_\theta \subset \mathscr{M}_\theta$ and, therefore,

$$l_{f_0,\theta} = \sup_{P_\theta \in \widehat{\mathscr{M}}_\theta} P_\theta\left[-\bar{f}_0\right] = -\inf_{P_\theta \in \widehat{\mathscr{M}}_\theta} P_\theta\left[\bar{f}_0\right] \leqslant -\underline{P}_\theta\left[\bar{f}_0\right] \leqslant -\underline{P}_\theta[f_0] \qquad \square$$

3.2. Procedure

Recall the notation from the previous subsection and assume that $(\overline{P}_\theta)_{\theta\in\Theta}$ is an imprecise model on the sample space $(\mathscr{X},\mathscr{A})$ such that (11)–(13) are fulfilled. Let $f_1,\ldots,f_n$ be elements of $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ such that

$$\{f_1,\ldots,f_n\} = \mathscr{K} = \bigcup_{\theta\in\Theta} \mathscr{K}_\theta$$

and define $\mathscr{I}_\theta := \{i \in \{1,\ldots,n\} | \ \ f_i \in \mathscr{K}_\theta\} \ \ \forall \theta \in \Theta$. Proceed in the following way for any fixed $\varepsilon \in (0,1)$:

STEP 1: For every $i \in \{1,\ldots,n\}$, take $d_i = d_{f_i}$ from (13) and

$$\varepsilon_i := \frac{\sup f_i - \inf f_i}{c \cdot d_i} \cdot \varepsilon \quad \text{where} \quad c := \sup_{\theta\in\Theta} \sum_{j \in \mathscr{I}_\theta} \frac{\sup f_j - \inf f_j}{d_j} \tag{17}$$

Note that the validity of

$$\sum_{j\in\{1,\ldots,n\}} \frac{\sup f_j - \inf f_j}{d_j} \geqslant c \geqslant \frac{\sup f_i - \inf f_i}{d_i}$$

ensures $0 < \varepsilon_i \leqslant \varepsilon < 1$. There is an $M \in \mathbb{N}$ such that

$$M - 1 \leqslant \frac{c}{\varepsilon} \leqslant M \tag{18}$$

For every $i \in \{1,\ldots,n\}$, define

$$b_i^{(j)} := \inf f_i + \frac{j}{M}(\sup f_i - \inf f_i) \quad \forall j \in \{0,1,2,\ldots,M\}$$

and $C_i^{(1)} := f_i^{-1}\left(\left[b_i^{(0)}, b_i^{(1)}\right]\right)$ and

$$C_i^{(j)} := f_i^{-1}\left(\left(b_i^{(j-1)}, b_i^{(j)}\right]\right) \in \mathscr{A} \quad \forall j \in \{2,\ldots,M\} \tag{19}$$

Then, define

$$s_i := \sum_{j=1}^{M} b_i^{(j)} I_{C_i^{(j)}} \quad \text{and note that} \quad f_i \leqslant s_i \leqslant f_i + \varepsilon_i d_i \tag{20}$$

Let $\mathscr{C}$ be the smallest σ-algebra which contains $C_i^{(j)}$ for every $j \in \{1,\ldots,M\}$ and every $i \in \{1,\ldots,n\}$. Since $\mathscr{C}$ is finite, there is a finite partition $C_1,\ldots,C_r$ of $\mathscr{X}$ such that every element of $\mathscr{C}$ is the union of some elements of the partition $C_1,\ldots,C_r$.
STEP 2: For every $\theta \in \Theta$, let $\overline{Q}_\theta$ be the coherent upper prevision on $\mathscr{L}_\infty(\mathscr{X},\mathscr{C})$ which corresponds to the credal set

$$\mathscr{N}_\theta = \left\{ Q_\theta \in \mathrm{ba}_1^+(\mathscr{X},\mathscr{C}) \,\middle|\, Q_\theta[s_i] \leqslant \overline{P}_\theta[f_i] + \varepsilon_i d_i \quad \forall i \in \mathscr{I}_\theta \right\}$$

Values of $\overline{Q}_\theta$ can be calculated by linear programs. To this end, choose any $x_k \in C_k$ for every $k \in \{1,\ldots,r\}$. Then: For any $f \in \mathscr{L}_\infty(\mathscr{X},\mathscr{C})$, consider

$$(f(x_1),\ldots,f(x_r)) \cdot q \to \max_q$$

where

$$(s_i(x_1),\ldots,s_i(x_r)) \cdot q \leqslant \overline{P}_\theta[f_i] + \varepsilon_i d_i \quad \forall i \in \mathscr{I}_\theta$$

and

$$q \in \mathbb{R}^r, q_k \geqslant 0 \quad \forall k \in \{1,\ldots,r\}, q_1 + \ldots + q_r = 1$$

The optimal value of this linear program is equal to $\overline{Q}_\theta[f]$.
STEP 3: Instead of the original imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on the (infinite) sample space $(\mathscr{X},\mathscr{A})$, consider the imprecise model $(\overline{Q}_\theta)_{\theta\in\Theta}$ on the *finite* sample space $(\mathscr{X},\mathscr{C})$ and solve the corresponding decision problem.
The following notation is used:

Notation 3.2. Recall the setup presented in Section 1.2 where we have an index set Θ, a decision space $(\mathbb{D},\mathscr{D})$, a loss function W and an imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on a sample space $(\mathscr{X},\mathscr{A})$. The task is to find an $\mathscr{A}$-measurable (randomized) decision function τ which minimizes the risk function $R((\overline{P}_\theta)_{\theta\in\Theta},\tau,W)$. This decision problem is called *original decision problem*.

Let the index set Θ, the decision space $(\mathbb{D},\mathscr{D})$ and the loss function W remain unchanged. But, now, let the imprecise model be $(\overline{Q}_\theta)_{\theta\in\Theta}$ on the *finite* sample space $(\mathscr{X},\mathscr{C})$ where $(\overline{Q}_\theta)_{\theta\in\Theta}$ and $\mathscr{C}$ are constructed by the above discretization procedure. The task is to find a $\mathscr{C}$-measurable (randomized) decision function κ which minimizes the risk function $R\Big((\overline{Q}_\theta)_{\theta\in\Theta},\kappa,W\Big)$. This decision problem is called (ε–)*discretized decision problem*.

Analogously to $\mathscr{T}_d(\mathscr{A})$ and $\mathscr{T}_0(\mathscr{A})$, the symbols $\mathscr{T}_0(\mathscr{C})$ and $\mathscr{T}_d(\mathscr{C})$ denote the $\mathscr{C}$-measurable (randomized) decision functions, respectively.

3.3. Correctness

In the present subsection, it is shown that a decision problem can be approximately solved by solving the corresponding discretized decision problem. To this end, recall the setup presented in Section 1.2 and Notation 3.2. In particular, this means that $(\overline{P}_\theta)_{\theta\in\Theta}$ is an imprecise model on the sample space $(\mathscr{X},\mathscr{A})$ and $(\overline{Q}_\theta)_{\theta\in\Theta}$ is the corresponding discretized model on the discretized sample space $(\mathscr{X},\mathscr{C})$ according to Section 3.2. For every $\theta \in \Theta$, let $\mathscr{M}_\theta$ be the credal set of $\overline{P}_\theta$, $\mathscr{N}_\theta$ be the credal set of $\overline{Q}_\theta$ on $(\mathscr{X},\mathscr{C})$ and $\overline{Q}'_\theta$ be the natural extension of $\overline{Q}_\theta$ on $\mathscr{L}_\infty(\mathscr{X},\mathscr{A})$ with credal set $\mathscr{N}'_\theta \subset \mathrm{ba}_1^+(\mathscr{X},\mathscr{A})$.

The proof consists of two steps: firstly, it has to be shown that replacing $(\overline{P}_\theta)_{\theta\in\Theta}$ by $(\overline{Q}'_\theta)_{\theta\in\Theta}$ only has small effects. Secondly, it has to be shown that, in case of model $(\overline{Q}'_\theta)_{\theta\in\Theta}$, reducing the (infinite) sample space from $(\mathscr{X},\mathscr{A})$ to the (finite) sample space $(\mathscr{X},\mathscr{C})$ does not have any effect. This is illustrated in the following diagram:

$$\begin{array}{lll}
(\mathscr{X},\mathscr{C}) & & (\overline{Q}_\theta)_{\theta\in\Theta} \\
 & & \uparrow \text{Lemma 3.4} \\
(\mathscr{X},\mathscr{A}) & (\overline{P}_\theta)_{\theta\in\Theta} \overset{\text{Lemma 3.3}}{\longrightarrow} & (\overline{Q}'_\theta)_{\theta\in\Theta}
\end{array}$$

Lemma 3.3 states that the credal sets of $\overline{P}_\theta$ and $\overline{Q}'_\theta$ approximately coincide. This is a consequence of the results of Section 2 about stable natural extensions.

Lemma 3.3. *Assume that the imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on $(\mathscr{X},\mathscr{A})$ fulfills (11)–(13). Then, for every $\theta \in \Theta : \mathscr{M}_\theta \subset \mathscr{N}_\theta$ and, for every $Q'_\theta \in \mathscr{N}'_\theta$, there is a $P_\theta \in \mathscr{M}_\theta$ such that*

$$\left| P_\theta[f] - Q'_\theta[f] \right| \leqslant \varepsilon(\sup f - \inf f) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X},\mathscr{A}) \tag{21}$$

Proof. Fix any $\theta \in \Theta$ and recall the definitions from Section 3.2.

For every $P_\theta \in \mathscr{M}_\theta$, the definition of s_i implies

$$P_\theta[s_i] \;\leqslant\; P_\theta[f_i + \varepsilon_i d_i] = P_\theta[f_i] + \varepsilon_i d_i \leqslant \overline{P}_\theta[f_i] + \varepsilon_i d_i \quad \forall i \in \mathscr{I}_\theta$$

and, therefore, the definition of $\overline{Q}'_\theta$ implies $P_\theta \in \mathcal{N}'_\theta$. Hence, $\mathcal{M}_\theta \subset \mathcal{N}_\theta$ and

$$\overline{P}_\theta[f] \leqslant \overline{Q}'_\theta[f] \quad \forall f \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A}) \tag{22}$$

Next, consider the coherent upper prevision $\overline{P}'_\theta$ on $\mathcal{L}_\infty(\mathcal{X},\mathcal{A})$ defined by

$$\overline{P}'_\theta[f] = \sup\left\{P'_\theta[f] \,\middle|\, P'_\theta \in \mathrm{ba}_1^+(\mathcal{X},\mathcal{A})),\ \ P'_\theta[f_i] \leqslant \overline{Q}'_\theta[f_i] \forall i \in \mathcal{I}_\theta\right\}$$

for every $f \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A})$. Together with (22), we have

$$\overline{P}_\theta[f] \leqslant \overline{Q}'_\theta[f] \leqslant \overline{P}'_\theta[f] \quad \forall f \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A}) \tag{23}$$

and

$$\overline{P}'_\theta[f_i] = \overline{Q}'_\theta[f_i] \quad \forall i \in \mathcal{I}_\theta \tag{24}$$

Then, it follows from the definition of d_i that

$$\overline{P}_\theta[f_i] \overset{(23)}{\leqslant} \overline{P}'_\theta[f_i] \overset{(24)}{=} \overline{Q}'_\theta[f_i] \leqslant \overline{Q}'_\theta[s_i] \leqslant \overline{P}_\theta[f_i] + \varepsilon_i d_i \leqslant \overline{P}_\theta[f_i] + \varepsilon_i\left(\overline{P}_\theta[f_i] - \underline{P}_\theta[f_i]\right) \quad \forall i \in \mathcal{I}_\theta$$

Let $\mathcal{M}'_\theta$ denote the credal set of $\overline{P}'_\theta$. The coherent upper prevision $\overline{P}'_\theta$ is the natural extension of a corresponding coherent upper prevision on $\mathcal{K}_\theta$ and the definitions ensure $\sum_{i\in\mathcal{I}_\theta}\varepsilon_i \leqslant \varepsilon$. Therefore an application of Lemma 2.3 for $\overline{P}_\theta$ and $\overline{P}'_\theta$ implies: For every $P'_\theta \in \mathcal{M}'_\theta$, there is a $P_\theta \in \mathcal{M}_\theta$ such that

$$\left|P_\theta[f] - P'_\theta[f]\right| \leqslant \varepsilon(\sup f - \inf f) \quad \forall f \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A}) \tag{25}$$

This implies (21) since $\mathcal{M}'_\theta \supset \mathcal{N}'_\theta$ according to (23). $\square$

Lemma 3.4. *Take $\star = 0$ or $\star = d$. Then, for every $\tau \in \mathcal{T}_\star(\mathcal{A})$, there is a $\kappa \in \mathcal{T}_\star(\mathcal{C})$ such that $R\left(\overline{Q}_\theta,\kappa,W\right) \leqslant R\left(\overline{Q}'_\theta,\tau,W\right)$ for every $\theta \in \Theta$.*

Proof. Take any $\tau \in \mathcal{T}_\star(\mathcal{A})$. Let $C_1,\ldots,C_r \in \mathcal{C}$ be the partition of $\mathcal{X}$ defined in STEP 1 of the discretization procedure in Section 3.2. For every $k \in \{1,\ldots,r\}$, choose any $x_k \in C_k$. For every $x \in \mathcal{X}$, let $\xi(x) = x_k$ if $x \in C_k$. This defines a $\mathcal{C}/\mathcal{A}$-measurable map $\xi : \mathcal{X} \to \mathcal{X}$, $x \mapsto \xi(x)$. Furthermore, define

$$\rho : \mathrm{ba}_1^+(\mathcal{X},\mathcal{C}) \to \mathrm{ba}_1^+(\mathcal{X},\mathcal{A}), \quad Q \mapsto \rho(Q)$$

where $\rho(Q)[f] = Q[f \circ \xi]$ for every $Q \in \mathrm{ba}_1^+(\mathcal{X},\mathcal{C})$ and $f \in \mathcal{L}_\infty(\mathcal{X},\mathcal{A})$. It is easy to see that this map is defined well; in particular ρ is a generalized randomization in the sense of [19]. Since $\rho(Q_\theta)$ is an extension to a finitely additive probability measure on $(\mathcal{X},\mathcal{A})$ for every $Q_\theta \in \mathcal{N}_\theta$, it follows that

$$\rho(\mathcal{N}_\theta) \subset \mathcal{N}'_\theta \tag{26}$$

In addition, it is easy to see that

$$\kappa : \mathcal{X} \times \mathcal{D} \to [0,1], \quad (x,D) \mapsto \kappa_x(D) = \tau_{\xi(x)}(D)$$

defines a $\mathcal{C}$-measurable (randomized) decision function $\kappa \in \mathcal{T}_\star(\mathcal{C})$. Next,

$$\sup_{Q_\theta\in\mathcal{N}_\theta} \int_{\mathcal{X}}\int_{\mathbb{D}} W_\theta(t)\,\tau_{\xi(x)}(dt)\,Q_\theta(dx) = \sup_{Q_\theta\in\mathcal{N}_\theta} \int_{\mathcal{X}}\int_{\mathbb{D}} W_\theta(t)\,\tau_x(dt)\,\rho(Q_\theta)(dx) \overset{(26)}{\leqslant} \sup_{Q'_\theta\in\mathcal{N}'_\theta} \int_{\mathcal{X}}\int_{\mathbb{D}} W_\theta(t)\,\tau_x(dt)\,Q'_\theta(dx)$$

implies $R\left(\overline{Q}_\theta,\kappa,W\right) \leqslant R\left(\overline{Q}'_\theta,\tau,W\right)$ for every $\theta \in \Theta$. $\square$

The theoretical properties of the discretization method with respect to risk functions are summarized in the following theorem.

Theorem 3.5. *Recall the setup presented in Section 1.2 and Notation 3.2. Assume that the imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on $(\mathcal{X},\mathcal{A})$ fulfills (11)–(13). Then:*

(a) For every $\mathcal{C}$-measurable randomized decision function $\kappa \in \mathcal{T}_0(\mathcal{C})$, the risk function with respect to the discretized decision problem is approximately equal to the risk function with respect to the original decision problem – more precisely:

$$R(\overline{P}_\theta,\kappa,W) \leqslant R\left(\overline{Q}_\theta,\kappa,W\right) \leqslant R(\overline{P}_\theta,\kappa,W) + \varepsilon(\sup W_\theta - \inf W_\theta)$$

for every $\theta \in \Theta$. In particular, the risk function with respect to the discretized decision problem is an upper bound for the risk function with respect to the original decision problem.

(b) Take $\star = 0$ or $\star = d$. Then, for every $\tau \in \mathcal{T}_\star(\mathcal{A})$, there is a $\kappa \in \mathcal{T}_\star(\mathcal{C})$ such that

$$R\left(\overline{Q}_\theta,\kappa,W\right) \leqslant R(\overline{P}_\theta,\tau,W) + \varepsilon(upW_\theta - \inf W_\theta) \quad \forall\theta \in \Theta$$

Informally speaking, part (a) states that, for $\mathscr{C}$-measurable (randomized) decision functions, it hardly matters if they are applied in the original or in the discretized decision problem. And part (b) states that the best $\mathscr{A}$-measurable (randomized) decision functions is hardly better than the best $\mathscr{C}$-measurable one.

Proof

(a) An application of Lemma 3.3 yields

$$\overline{P}_\theta[f] \leqslant \overline{Q}'_\theta[f] \leqslant \overline{P}_\theta[f] + \varepsilon(\sup f - \inf f) \quad \forall f \in \mathscr{L}_\infty(\mathscr{X}, \mathscr{A}) \tag{27}$$

for every $\theta \in \Theta$ and this implies part (a) according to the definition of the risk function.

(b) Take any $\tau \in \mathscr{T}_\star(\mathscr{A})$. Then, Lemma 3.4 implies existence of a $\kappa \in \mathscr{T}_\star(\mathscr{C})$ such that

$$R\left(\overline{Q}_\theta, \kappa, W\right) \leqslant R\left(\overline{Q}'_\theta, \tau, W\right)$$

for every $\theta \in \Theta$. Next, it follows from (27) that

$$R\left(\overline{Q}'_\theta, \tau, W\right) \leqslant R(\overline{P}_\theta, \tau, W) + \varepsilon(\sup W_\theta - \inf W_\theta)$$

for every $\theta \in \Theta$. Together, this implies part (b). $\square$

One of the most common optimality criterion for (randomized) decision functions is the so-called Γ-minimax criterion. Corollary 3.6 states that, in case of this optimality criterion, the original decision problem may in fact be approximately solved by the discretized decision problem. That is, an approximately optimal (randomized) decision function in the original decision problem can be found by searching for an optimal (randomized) decision function in the discretized decision problem.

In order to formulate this criterion, we need an (imprecise) prior distribution given by a coherent upper prevision

$$\overline{\Pi} : \mathscr{L}_\infty(\Theta, 2^\Theta) \to \mathbb{R},$$

and we assume that

$$\sup |W| < \infty \tag{28}$$

Let

$$\mathscr{P} = \left\{\pi \in \mathrm{ba}_1^+\left(\Theta, 2^\Theta\right) \middle|\ \pi[h] \leqslant \overline{\Pi}[h] \ \ \forall h \in \mathscr{L}_\infty\left(\Theta, 2^\Theta\right)\right\}$$

be the corresponding credal set where 2^Θ denotes the power set of Θ. With respect to the prior $\overline{\Pi}$, the (upper) Bayes risk of an $\mathscr{A}$-measurable (randomized) decision function τ in the original decision problem is defined to be

$$R_{\overline{\Pi}}((\overline{P}_\theta)_{\theta\in\Theta}, \tau, W) := \sup_{\pi \in \mathscr{P}} \int R(\overline{P}_\theta, \tau, W)\pi(d\theta)$$

Accordingly, the (upper) Bayes risk of a $\mathscr{C}$-measurable (randomized) decision function κ in the discretized decision problem is defined to be

$$R_{\overline{\Pi}}\left((\overline{Q}_\theta)_{\theta\in\Theta}, \kappa, W\right) := \sup_{\pi \in \mathscr{P}} \int R\left(\overline{Q}_\theta, \kappa, W\right)\pi(d\theta)$$

A (randomized) decision function is called Γ-minimax if it minimizes the respective upper Bayes risk over all (randomized) decision functions.

Corollary 3.6. *Recall the setup presented in Section 1.2 and Notation 3.2. Assume that the imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ on $(\mathscr{X}, \mathscr{A})$ fulfills (11)–(13) and assume that the loss function fulfills (28). Take $\star = 0$ or $\star = d$. Let $\tilde{\kappa} \in \mathscr{T}_\star(\mathscr{C})$ minimize the upper Bayes risk in the discretized decision problem, i.e.*

$$R_{\overline{\Pi}}\left((\overline{Q}_\theta)_{\theta\in\Theta}, \tilde{\kappa}, W\right) = \inf_{\kappa \in \mathscr{T}_\star(\mathscr{C})} R_{\overline{\Pi}}\left((\overline{Q}_\theta)_{\theta\in\Theta}, \kappa, W\right)$$

Then, $\tilde{\kappa} \in \mathscr{T}_\star(\mathscr{A})$ and $\tilde{\kappa}$ approximately minimizes the upper Bayes risk in the original decision problem, i.e.

$$R_{\overline{\Pi}}((\overline{P}_\theta)_{\theta\in\Theta}, \tilde{\kappa}, W) \leqslant \inf_{\tau \in \mathscr{T}_\star(\mathscr{A})} R_{\overline{\Pi}}((\overline{P}_\theta)_{\theta\in\Theta}, \tau, W) + \varepsilon(\sup W - \inf W)$$

Proof. Take any $\hat{\tau} \in \mathscr{T}_\star(\mathscr{A})$. Then, according to Theorem 3.5b), there is some $\hat{\kappa} \in \mathscr{T}_\star(\mathscr{C})$ such that

$$R\left(\overline{Q}_\theta, \hat{\kappa}, W\right) \leqslant R(\overline{P}_\theta, \hat{\tau}, W) + \varepsilon(\sup W_\theta - \inf W_\theta) \quad \forall \theta \in \Theta$$

Hence, the definition of the upper Bayes risk implies

$$R_{\overline{\Pi}}\left((\overline{Q}_\theta)_{\theta\in\Theta}, \hat{\kappa}, W\right) \leqslant R_{\overline{\Pi}}((\overline{P}_\theta)_{\theta\in\Theta}, \hat{\tau}, W) + \varepsilon(\sup W - \inf W) \tag{29}$$

By use of Theorem 3.5a), it follows that

$$R_{\overline{\Pi}}\big((\overline{P}_\theta)_{\theta\in\Theta},\tilde{\kappa},W\big) \leqslant R_{\overline{\Pi}}\Big((\overline{Q}_\theta)_{\theta\in\Theta},\tilde{\kappa},W\Big) \tag{30}$$

Assertions (29) and (30) and optimality of $\tilde{\kappa}$ (in the discretized decision problem) imply

$$R_{\overline{\Pi}}\big((\overline{P}_\theta)_{\theta\in\Theta},\tilde{\kappa},W\big) \leqslant R_{\overline{\Pi}}\big((\overline{P}_\theta)_{\theta\in\Theta},\hat{\tau},W\big) + \varepsilon(\sup W - \inf W)$$

This is true for every $\hat{\tau} \in \mathscr{T}_\star(\mathscr{C})$ and, therefore, the validity of Corollary 3.6 is assured. $\square$

Though [13] is concerned with discretizing Θ, the setup of the present section is closely related to the setup in Troffaes [13]. In the above described discretization method, the discrete sample space $(\mathscr{X},\mathscr{C})$ is generated by some simple functions s where every simple function s corresponds to some $f \in \mathscr{K}$ such that

$$\sup_{x\in\mathscr{X}} |f(x) - s(x)| = \max_{A\in\mathscr{A}} \sup_{x\in A} |f(x) - s(x)| \leqslant \varepsilon(\sup f - \inf f)$$

This is denoted by

$$f \sim_\varepsilon s$$

in Troffaes [13, Section 3]. Furthermore, $(\overline{Q}_\theta)_{\theta\in\Theta}$ is an imprecise model on $(\mathscr{X},\mathscr{C})$ where $(\mathscr{N}_\theta)_{\theta\in\Theta}$ is the corresponding family of credal sets on $(\mathscr{X},\mathscr{C})$. It follows from Lemma 3.3 that

$$\inf_{P_\theta\in\mathscr{M}_\theta} \|Q_\theta - P_\theta\| \leqslant 2\varepsilon \quad \forall Q_\theta \in \mathscr{N}_\theta$$

and

$$\inf_{Q_\theta\in\mathscr{N}_\theta} \|P_\theta - Q_\theta\| \leqslant 2\varepsilon \quad \forall P_\theta \in \mathscr{M}_\theta$$

for every $\theta \in \Theta$. This is denoted by

$$\mathscr{M}_\theta \sim_{2\varepsilon} \mathscr{N}_\theta$$

in Troffaes [13, Section 3]. Here, $\|\cdot\|$ denotes the total variation norm in $\mathrm{ba}_1^+(\mathscr{X},\mathscr{C})$; see [18, p. 240].

Furthermore, adopting the terminology from Troffaes [13, Section 4], Corollary 3.6 may be reformulated in the following way:

> Every (randomized) decision function on $(\mathscr{X},\mathscr{C})$ which is optimal in the ε-discretized decision problem is ε-optimal in the original decision problem.

Accordingly, Corollary 3.6 corresponds to [13, Theorem 6]. However, Corollary 3.6 is concerned with discretizing the sample space $(\mathscr{X},\mathscr{A})$ whereas [13, Theorem 6] is concerned with discretizing Θ.

4. Applicability

The above presented discretization method can be applied step by step. In particular, every value which has to be calculated can in principle be calculated by linear programming. However, rigid applications may in general be handicapped – or even made impossible – because of exceedingly high computational costs. This is again similar to the results in Troffaes [13] and we may derive upper bounds for the size of the discretized sample space which generally holds but which are, in general, much too large in order to be of any practical value.

As already stated in Section 3.2, there is a finite partition $\{C_1,\ldots,C_r\}$ of $\mathscr{C}$ such that every element of $\mathscr{C}$ is the union of some elements of the partition $\{C_1,\ldots,C_r\}$. The size of this partition – i.e. the number $r \in \mathbb{N}$ – precisely corresponds to the size of the discretized sample space: r is the number of possible (discrete) observations after discretizing.

According to the definition of $\mathscr{C}$, the partition $\{C_1,\ldots,C_r\}$ is the coarsest partition which is finer than every partition

$$\{C_i^{(1)},\ldots,C_i^{(M)}\}, \quad i \in \{1,\ldots,n\}$$

where $C_i^{(j)}$ is defined in (19) for every $i \in \{1,\ldots,n\}$ and $j \in \{1,\ldots,M\}$. Therefore, an upper bound on r is given by

$$r \leqslant M^n \leqslant \left(1 + \frac{1}{\varepsilon}\cdot\sup_{\theta\in\Theta}\sum_{j\in\mathscr{I}_\theta}\frac{\sup f_j - \inf f_j}{d_j}\right)^n \tag{31}$$

where the last inequality follows from (17) and (18). This number is extremely large – even if Θ is a small set and, for every $\theta \in \Theta$, $\mathscr{K}_\theta$ only contains a few elements. For example, let Θ contain 10 elements, and, for every $\theta \in \Theta$ let each $\mathscr{K}_\theta$ also contain 10 elements such that $\mathscr{K}_{\theta_1},\ldots,\mathscr{K}_{\theta_{10}}$ are pairwise disjoint. Therefore, we have $n = 100$. Furthermore, assume for simplicity that

$$\overline{P}_\theta[f] - \underline{P}_\theta[f] = 0.1\cdot(\sup f - \inf f) \quad \forall f \in \mathscr{K}_\theta \quad \forall \theta \in \Theta$$

Then, for $\varepsilon = 0.1$, the number in (31) is

$$\left(1 + \frac{1}{0.1} \cdot 10 \cdot \frac{1}{0.1}\right)^{100} > 10^{300}$$

However, this number usually decreases immensely: it is unrealistic to assume that $\mathscr{K}_{\theta_1}, \ldots, \mathscr{K}_{\theta_{10}}$ are pairwise disjoint in applications. In most applications, $\mathscr{K}_\theta$ will not depend on θ so that we have

$$\mathscr{K} = \mathscr{K}_\theta \quad \forall \theta \in \Theta$$

In this case, n does not increase with the number of elements of Θ and we would get $n = 10$ in the above example. This leads to the number

$$\left(1 + \frac{1}{0.1} \cdot 10 \cdot \frac{1}{0.1}\right)^{10} \approx 10^{30}$$

which still is a great deal too large. However, (31) only is a very crude upper bound which does not assume any additional properties of the functions $f \in \mathscr{K}$. Such assumptions may drastically decrease the bound as can be seen by Proposition 4.1.

Proposition 4.1. *Let $\mathscr{X}$ be an interval in $\mathbb{R}$ and assume that every $f \in \mathscr{K}$ fulfills one of the following properties:*

(a) f is the indicator function of a set $A \in \mathscr{A}$ which is the union of no more than

$$\frac{1}{\varepsilon} \cdot \sup_{\theta \in \Theta} \sum_{j \in \mathscr{I}_\theta} \frac{\sup f_j - \inf f_j}{d_j}$$

intervals.
(b) f is convex.
(c) f is concave.

Let $\{C_1, \ldots, C_r\}$ be the partition from Section 3.2. Then,

$$r \leqslant 2n \cdot \left(1 + \frac{1}{\varepsilon} \cdot \sup_{\theta \in \Theta} \sum_{j \in \mathscr{I}_\theta} \frac{\sup f_j - \inf f_j}{d_j}\right) - n \tag{32}$$

Roughly speaking, property (a) means that f is the indicator function of a set $A \in \mathscr{A}$ which is not too complicated.

Proof. Recall that $\mathscr{K} = \{f_1, \ldots, f_n\}$ and recall that $\{C_i^{(1)}, \ldots, C_i^{(M)}\}$ is the partition defined by (19) for every $i \in \{1, \ldots, n\}$.
Firstly, it is shown for every $i \in \{1, \ldots, n\}$ that there is another partition $\{B_i^{(1)}, \ldots, B_i^{(2M)}\}$ of $\mathscr{X}$ such that

- $B_i^{(j)}$ is an interval in $\mathbb{R}$ for every $j \in \{1, \ldots, 2M\}$ and
- $\{C_i^{(1)}, \ldots, C_i^{(M)}\}$ is contained in the smallest σ-algebra generated by $\{B_i^{(1)}, \ldots, B_i^{(2M)}\}$.

For any $i \in \{1, \ldots, n\}$ such that f_i fulfills (a), this follows immediately from (18). Now, take any $i \in \{1, \ldots, n\}$ such that f_i fulfills (b). Then, the definition of $C_i^{(j)}$ and convexity of f_i implies that $C_i^{(j)}$ is the union of two intervals $B_i^{(j)}$ and $B_i^{(2j)}$ for every $j \in \{1, \ldots, M\}$. The same is true in case of (c).
Next, note that the number of elements of each partition $\{B_i^{(1)}, \ldots, B_i^{(2M)}\}$ is bounded by

$$2M \leqslant 2 \cdot \left(1 + \frac{1}{\varepsilon} \cdot \sup_{\theta \in \Theta} \sum_{j \in \mathscr{I}_\theta} \frac{\sup f_j - \inf f_j}{d_j}\right)$$

Finally, Proposition 4.1 follows from the following simple fact: Let $D_1^{(1)}, \ldots, D_1^{(m_1)}$ and $D_2^{(1)}, \ldots, D_2^{(m_2)}$ be two partitions of an interval in $\mathbb{R}$ such that every $D_i^{(j)}$ is an interval. Let $l_i^{(j)}$ be the corresponding left endpoints of these intervals. Without loss of generality, assume that $l_i^{(1)} \leqslant l_i^{(2)} \leqslant \ldots \leqslant l_i^{(m_i)}$. Then, there is a common refinement $D_1, \ldots, D_{r'}$ (consisting of intervals again) such that, for every $k \in \{1, \ldots, r'\}$, the left endpoint of D_k is an element of $\{l_1^{(1)}, \ldots, l_1^{(m_1)}, l_2^{(1)}, \ldots, l_2^{(m_2)}\}$. Since $l_1^{(1)} = l_2^{(1)}$, this implies the following bound on the size of the common refinement: $r' \leqslant m_1 + m_2 - 1$. $\square$

In the situation of the above example with $n = 10$, this leads to the upper bound

$$2 \cdot 10 \cdot \left(1 + \frac{1}{0.1} \cdot 10 \cdot \frac{1}{0.1}\right) - 10 \approx 2 \cdot 10^4$$

which is a more reasonable size than the above ones. In particular, bound (32) has the remarkable property that it increases only linearly(!) in n, the number of functions. On the one hand, Proposition 4.1 itself covers many situations in real applications. On the other hand, it demonstrates, that applying the presented discretization procedure will often lead to a reasonable size r of the discretized sample space.

Furthermore, there is another way to reduce the size of the discretized sample space $(\mathscr{X}, \mathscr{C})$, which is different from the others and relates to the results of Section 2. In the presented discretization method, an imprecise model $(\overline{Q}_\theta)_{\theta\in\Theta}$ is constructed such that

$$\overline{P}_\theta[f_i] \leqslant \overline{Q}_\theta[f_i] \leqslant \overline{P}_\theta[f_i] + \frac{\varepsilon}{c}(\sup f_i - \inf f_i) \quad \forall i \in \{1,\ldots,n\} \tag{33}$$

However, c can be large and then, it will not be possible in most applications to specify the "correct" coherent upper prevision in such a great precision that $\frac{\varepsilon}{c}(\sup f_i - \inf f_i)$ becomes meaningful in (33). Therefore, it may be justified to relax (33) to

$$\overline{P}_\theta[f_i] \leqslant \overline{Q}_\theta[f_i] \leqslant \overline{P}_\theta[f_i] + \varepsilon(\sup f_i - \inf f_i) \quad \forall i \in \{1,\ldots,n\} \tag{34}$$

This means, that M is not chosen in order to fulfill (18) in the discretization method. Instead, M has to be chosen so that

$$M - 1 < \frac{1}{\varepsilon} \leqslant M$$

Then, analog to (32), an upper bound on the size r would be

$$2 \cdot n \cdot \left(1 + \frac{1}{\varepsilon}\right) - n \tag{35}$$

and the above example would lead to

$$2 \cdot 10 \cdot \left(1 + \frac{1}{0.1}\right) - 10 = 210$$

This is a reasonable size with which computations should be tractable. Note that bound (35) does not depend on the size of Θ and only depends linearly on the number of elements in $\mathscr{K}$. Therefore, also larger problem than the above example should be tractable.

Such an alternative approach is often justified the more so as a large c indicates that the imprecise model $(\overline{P}_\theta)_{\theta\in\Theta}$ is in danger of being instable – cf. Section 2. Then, relaxing (33) to (34) corresponds to a more conservative approach. If this has a large effect on the results, this means that small changes of $\overline{P}_\theta[f_0]$, $f_0 \in \mathscr{K}_0$, have large effects on $\overline{P}_\theta[f]$ for some $f \notin \mathscr{K}_0$. In this unstable case, it seems to be a good idea to be more conservative because this may save from arbitrary results. However, note that, by doing this, ε-optimality is not guaranteed anymore according to the results in Section 3.3.

5. Conclusions

The present paper is concerned with discretizing sample spaces in data-based decision theory under imprecise probabilities. In particular, a method of discretizing is presented which can be executed step by step in real applications. In this way, the original decision problem is turned into a discretized decision problem which is accessible by linear programming. It is proven that a solution of the discretized problem approximately solves the original decision problem. Furthermore, it is shown that the discretized sample space keeps reasonably small in many applications.

Since algorithms which can deal with finite sets are already available, the results of the paper makes it possible to apply these algorithms also in case of infinite sets. This is, for example, possible in case of statistical hypothesis testing where [21] provides algorithms for calculating optimal tests under imprecise probabilities under the assumption that the sample space is finite. Now, these algorithms can, in principle, also be used in applications where the original sample space is infinite. However, this application also points out the limitation of the proposed method of discretizing: in mathematical statistics, the data $x_1,\ldots,x_n$ are usually treated as one single data point $x = (x_1,\ldots,x_n)$ by going over to product spaces. Though this is also possible in case of imprecise probabilities, large numbers of observations will lead to problems: Firstly, it is not clear in how far going over to product probabilities complies with Assumption (11). Secondly, the size of discrete problems (and, therefore, also the size of discretized problems) usually drastically increases with the number of observations. [2]

The present paper also points out that natural extensions are potentially most instable so that natural extensions are in danger of providing arbitrary results in applications. Due to the popularity of natural extensions in imprecise probabilities, this is an important topic which should attract more attention. In particular, it would be desirable to develop guidelines which safeguard from these instabilities. Theorem 2.2 is a first attempt into this direction which, at least, applies for the presented discretizing method.

Acknowledgement

I would like to thank Thomas Augustin for valuable suggestions. Furthermore, I am much indebted to the Cusanuswerk (Foundation of the Roman Catholic Church) which supported parts of this work by a Ph.D. scholarship. I thank the

[2] For a discussion of the latter problem in case of statistical hypothesis testing, see [21, Sections 4.1.4 and 6.1.2]. At least in case of hypothesis testing, this problem can sometimes be avoided: In the presence of so-called globally least favorable pairs, the complexity of testing problems does not increase with the number of observations; cf. [22, Satz II.B.4] and [23, Satz 2.57]. Least favorable pairs have attracted much attention after the publication of [24].

reviewers for helpful comments; in particular, a conjecture of one reviewer has led to an improvement of bound (32) in Proposition 4.1.

References

[1] A. Wald, Statistical Decision Functions, John Wiley & Sons Inc., New York, 1950.
[2] J. Berger, Statistical Decision Theory and Bayesian Analysis, second ed., Springer-Verlag, New York, 1985.
[3] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York, 1986.
[4] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman & Hall, London, 1991.
[5] K. Weichselberger, The theory of interval-probability as a unifying concept for uncertainty, International Journal of Approximate Reasoning 24 (2000) 149–170. reasoning with imprecise probabilities (Ghent, 1999.
[6] I. Gilboa, D. Schmeidler, Maxmin expected utility with non-unique prior, Journal of Mathematical Economics 18 (2) (1989) 141–153.
[7] M. Schervish, T. Seidenfeld, J. Kadane, I. Levi, Extensions of expected utility theory and some limitations of pairwise comparisons, in: J. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, Lugano, Carleton Scientific, Waterloo, 2003, pp. 496–510.
[8] M. Troffaes, Decision making under uncertainty using imprecise probabilities, International Journal of Approximate Reasoning 45 (2007) 17–29.
[9] P. Walley, R. Pelessoni, P. Vicig, Direct algorithms for checking consistency and making inferences from conditional probability assessments, Journal of Statistical Planning and Inference 126 (1) (2004) 119–151.
[10] C. de Campos, F. Cozman, Inference in credal networks through integer programming, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, Prague, 2007, pp. 145–154.
[11] T. Augustin, Optimal decisions under complex uncertainty – basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability, ZAMM. Zeitschrift für Angewandte Mathematik und Mechanik, Journal of Applied Mathematics and Mechanics 84 (10–11) (2004) 678–687.
[12] L. Utkin, T. Augustin, Powerful algorithms for decision making under partial prior information and general ambiguity attitudes, in: F. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA'05, Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, Pittsburg, SIPTA, Manno, 2005, pp. 349–358.
[13] M. Troffaes, Finite approximations to coherent choice, International Journal of Approximate Reasoning 50 (2009) 655–665.
[14] M. Obermeier, T. Augustin, Lucenos discretization methods and its application in decision making under ambiguity, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, Prague, 2007, pp. 327–336.
[15] T. Augustin, On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view: a cautionary note on updating imprecise priors, in: J. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA'03, Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, Lugano, Carleton Scientific, Waterloo, 2003, pp. 31–45.
[16] E. Miranda, A survey of the theory of coherent lower previsions, International Journal of Approximate Reasoning 48 (2008) 628–658.
[17] K.P.S. Bhaskara Rao, M. Bhaskara Rao, Theory of Charges, Academic Press Inc., New York, 1983.
[18] N. Dunford, J. Schwartz, Linear Operators. I. General Theory, Wiley-Interscience Publishers, New York, 1958.
[19] R. Hable, Data-based decisions under imprecise probability and least favorable models, International Journal of Approximate Reasoning 50 (2009) 642–654.
[20] R. Hable, Data-based decisions under complex uncertainty, Ph.D. Thesis. Ludwig-Maximilians-Universität (LMU) Munich, 2009, <http://edoc.ub.uni-muenchen.de/9874/>.
[21] T. Augustin, Optimale Tests bei Intervallwahrscheinlichkeit, Vandenhoeck & Ruprecht, Göttingen, 1998.
[22] H. Rieder, Robuste Tests, Diplomarbeit, Albert-Ludwigs-Universität Freiburg, 1974.
[23] H. Witting, Mathematische Statistik. I, B.G. Teubner, Stuttgart, 1985.
[24] P. Huber, V. Strassen, Minimax tests and the Neyman-Pearson lemma for capacities, The Annals of Statistics 1 (1973) 251–263.